

MC meetodid, praktikum 6

Teema 1: Üldise valikumeetodi rakendamine

Algoritm: Olgu f ja g tihedusfunktsioonidega f_X ja g_X proportsionaalsed funktsioonid ja c selline reaalarv, et kehtib tingimus $f(x) \leq cg(x) \forall x$.

1. Tekitame tihedusele g_X vastavad pseudojuhuslikud arvud $y_{11}, y_{12}, \dots, y_{1n}$
2. Genereerime n pseudojuhuslikku arvu $y_{21}, \dots, y_{2n}$ ühtlasest jaotusest $U(0, 1)$
3. Väljastame sellised y_{1i} väärtused, mille korral kehtib $y_{2i} \leq \frac{f(y_{1i})}{c \cdot g(y_{1i})}$

Tulemuseks saame tihedusfunktsioonile f_X vastavad pseudojuhuslikud arvud.

Harjutus 1

Kasuta 100000 sobiva parameetriga eksponentjaotusele vastavat pseudojuhuslikku arvu selleks, et genereerida tihedusfunktsioonile

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{8 \cos(x)^2}{3 \cdot e^{2x}}, & x \geq 0 \end{cases}$$

vastavaid pseudojuhuslikke arve. Kas saadud juhuslike suuruste arv on kooskõlas meetodi teoreetilise efektiivsusega? Leia nende põhjal hinnang tihedusfunktsiooniga f juhusliku suuruse standardhälbele.

Vihjed Kasuks tulevad järgmised käsud: **rexp**, **dexp** ja **sd**. Samuti saab soovi korral tulemust kontrollida, leides käsu **integrate** abil dispersiooni ja selle kaudu standardhälbe väärtuse numbrilise integreerimise teel.

Kontrolliks: tekkima peaks vähemalt 74000 väärtust soovitud jaotusest, kui c on valitud optimaalne. Standardhälbe hinnangu ja täpse väärtuse algus peaks olema 0.3...

Harjutus 2

Vaatleme tihedusfunktsioonile

$$f(x) = \begin{cases} \frac{\pi \sin(\pi x) + 4x}{6}, & 0 \leq x < 1, \\ (x-2)^2, & 1 \leq x < 2, \\ 0 & \text{mujal} \end{cases}$$

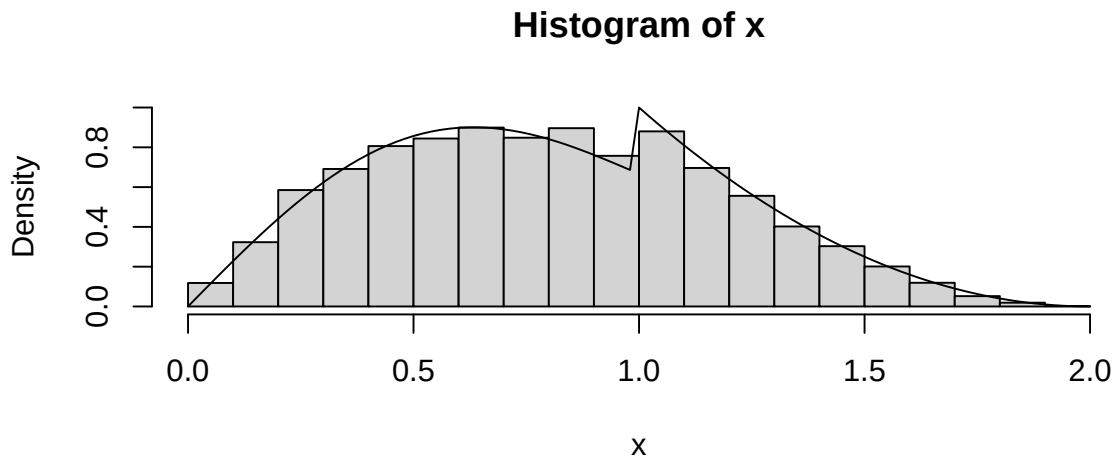
vastavate pseudojuhuslike arvude genereerimist üldise valikumeetodiga juhul, kus suuruseid y_1 genereeritakse jaotusfunktsioonile

$$G(x) = \begin{cases} 0, & x \leq 0, \\ \frac{x^2}{2}, & 0 < x \leq 1 \\ 1 - \frac{(2-x)^2}{2}, & 1 < x \leq 2 \\ 1, & x > 2 \end{cases}$$

vastavast jaotusest. Vali efektiivsusest lähtuvalt selline genereeritavate y_1 väärtuste arv, et tulemusena saaksime soovitud jaotusest ligikaudu 10000 pseudojuhuslikku arvu ning võrdle nende abil tekitatud histogrammi tihedusfunktsiooniga f .

Vihje: Kõigepealt tuleb leida funktsioonile G vastav tihedusfunktsioon; jaotusfunktsioonile G vastavate väärtuste saamiseks kasuta pöördfunktsiooni meetodit. Histogrammi joonistamisel tuleks kasutada suvandit `probability=TRUE`, samuti saab soovi korral tulpade arvu kontrollida suvandiga `breaks=m`, kus m on soovitud tulpade arv.

Kontrolliks: üldise valikumeetodi efektiivsus vaadeldava ülesande lahendamisel võiks olla suurem kui 0.4; tulemus võiks välja näha nagu allpool toodud pildil:



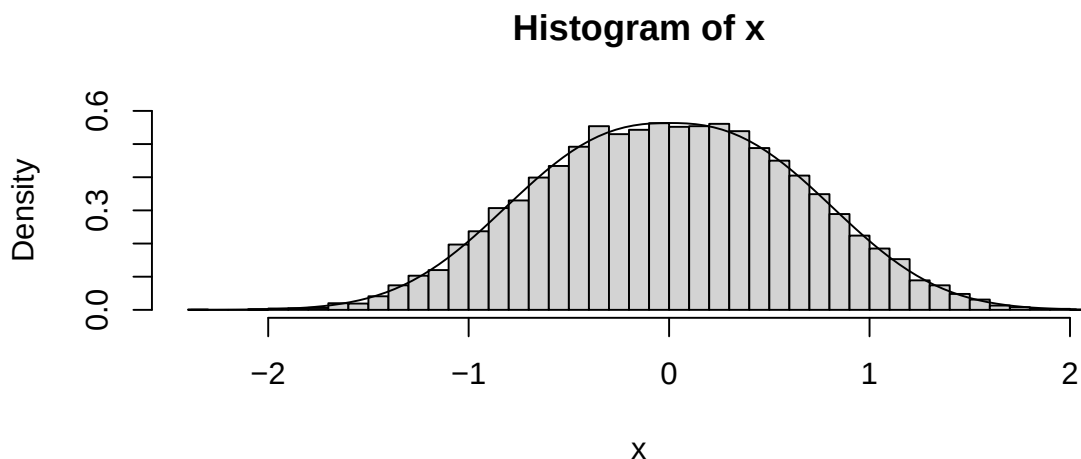
Üldist valikumeetodit võib rakendada ka siis, kui funktsioonide f ja g asemel kasutada vastavate tihedusfunktsioonidega proportsionaalseid funktsioone, kusjuures konstant c on ikka leitud tingimusest $f(x) \leq c \cdot g(x) \forall x$. Ainult peab teadma, et siis efektiivsus ei ole enam $\frac{1}{c}$.

Harjutus 3

Genereeri 10000 funktsiooniga $f(x) = e^{-|x|^{2.5}}$ proportsionaalse tihedusfunktsiooniga jaotusele vastavat pseudojuhuslikku arvu, kasutades selleks normaalfaotust, mille tihedusfunktsioon on proportsionaalne funktsiooniga $g(x) = e^{-x^2}$. Mis on saadud meetodi efektiivsuseks? Veenduda histogrammi abil, et saadud arvud käituvad vastavalt soovitud jaotusele.

Vihje: Et saada 10000 väärtust, tuleks genereerimisel kasutada sellist n väärtust, et üsna kindlalt tuleks rohem kui 10000 väärtust ja jätta neist alles nõutud arvu väärtuseid. Histogrammi jaoks vajaliku normeeriva konstandi saab leida käsku `integrate` kasutades.

Kontrolliks veendu, et valikumeetodi abil saadud väärtuste arv on vähemalt 10000; valikumeetodi efektiivsus võiks olla vähemalt 0.9, saadud graafik peaks sarnanema alltoodule.



Teema 2: Genereerimine segujaotusest

Sageli võib andmete jaotusi uurides täheldada mitut “küüru” või lihtsalt erinevat tüüpi käitumist erinevates väärtuste piirkondades, mistõttu ei vasta nende käitumine lihtsatele tuntud jaotustele. Tihti on selline käitumine põhjustatud üldkogumi mittehomoogeensusest ehk sellest, et üldkogumis on mitu erinevat rühma (näiteks mehed/naised), milles mõõdetav tunnus käitub erinevalt. Olgu meil m rühma, kusjuures igas rühmas käitugu uuritav tunnus vastavalt pidevale juhuslikule suurusele tihedusfunktsiooniga f_i , $i = 1, 2, \dots, m$. Olgu iga rühma osakaal üldkogumis p_i , siis täistõenäosuse valemi abil on lihtne näidata, et uuritava tunnuse tihedusfunktsioon üldkogumist juhusliku liikme valikul avaldub kujul

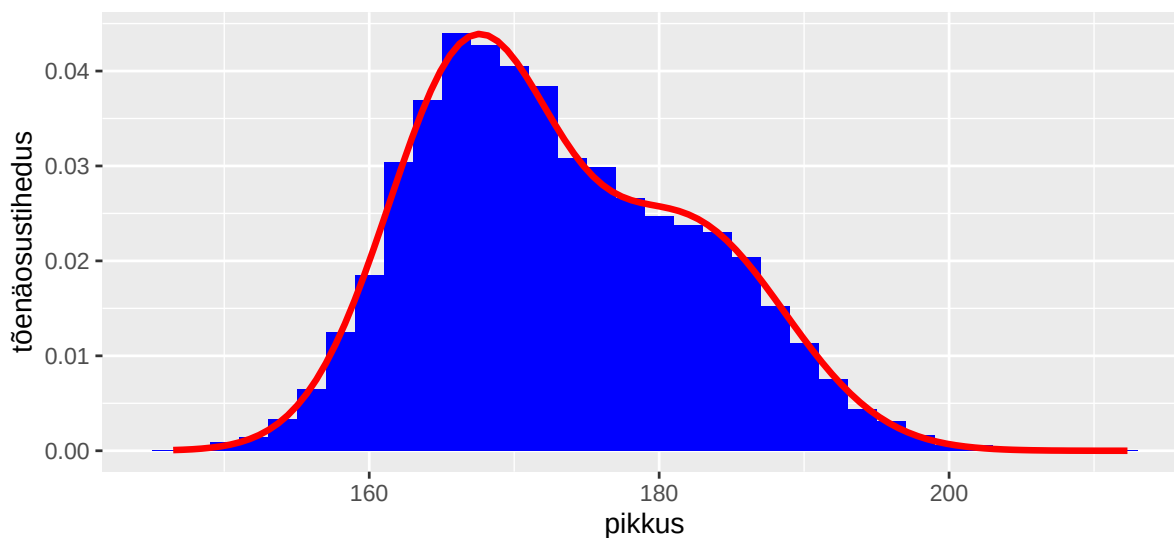
$$f(x) = \sum_{i=1}^m p_i f_i(x), \quad -\infty < x < \infty.$$

See valem annab meile võimaluse keerulisemalt käituvaid juhuslikke suuruseid esitada lihtsamate kaudu.

Harjutus 4

Oletame, et naissoost tudengite pikkused vastavad normaaljaotusele $N(167, \sqrt{33})$ ning meessoost tudengite pikkused vastavad normaaljaotusele $N(182, \sqrt{46})$. Kirjuta funktsioon, mis etteantud n ja meeste osakaalu p korral genereerib n juhuslikult valitud tudengi pikkust. Kujuta graafikul 16000 genereeritud pikkuse histogramm juhul $p = 0.4$ koos teoreetilise tihedusfunktsiooniga.

Kontrolliks: tulemus võiks näha välja nagu allpooltoodud graafik



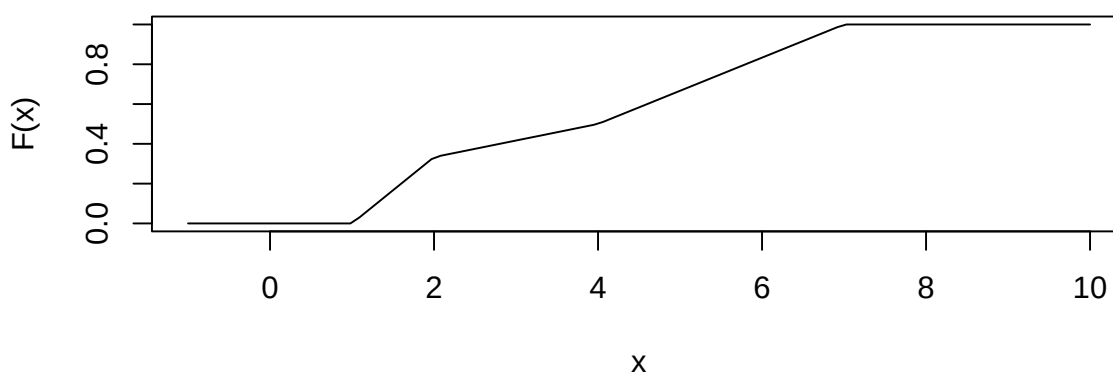
Harjutus 5

Kasutada segujaotuse ideed selleks, et genereerida 2000 pseudojuhuslikku arvu tihedusfunktsioonile

$$f(x) = \begin{cases} 0, & x < 1, \\ \frac{1}{3}, & 1 \leq x < 2, \\ \frac{1}{12}, & 2 \leq x < 4, \\ \frac{1}{6}, & 4 \leq x < 7, \\ 0, & x \geq 7 \end{cases}$$

vastavast jaotusest. Kontrollida Kolmogorov-Smirnovi testiga saadud arvude vastavust soovitud jaotusele.

Kontrolliks: leitud jaotusfunktsiooni graafik peaks olema selline nagu järgneval joonisel



Kodutöö nr 5 (1,5p, tähtaeg 16.10.2022)

Moodle VPL (Virtual Programming Lab) harjutus “Kodu5”.

Ülesande tekst on leitav Moodle keskkonnas. Lahendus esitada R skriptina (.R fail). Eelnevalt kontrollige, et lahendus töötab nõ puhtalt lehelt alustades (RStudios valida eelnevalt Session -> Clear Wokspace). Palun jälgige hoolikalt ülesande juhiseid!